The uselessness of AI ethics

Abstract

As the awareness of AI's power and danger has risen, the dominant response has been a turn to ethical principles. A flood of AI guidelines and codes of ethics have been released in both the public and private sector in the last several years. However, these are meaningless principles which are contested or incoherent, making them difficult to apply; they are isolated principles situated in an industry and education system which largely ignores ethics; and they are toothless principles which lack consequences and adhere to corporate agendas. For these reasons, I argue that AI ethical principles are useless, failing to mitigate the racial, social, and environmental damages of AI technologies in any meaningful sense. The result is a gap between high-minded principles and technological practice. Even when this gap is acknowledged and principles seek to be "operationalized," the translation from complex social concepts to technical rulesets is non-trivial. In a zero-sum world, the dominant turn to AI principles is not just fruitless but a dangerous distraction, diverting immense financial and human resources away from potentially more effective activity. I conclude by highlighting alternative approaches to AI justice that go beyond ethical principles: thinking more broadly about systems of oppression and more narrowly about accuracy and auditing.